Stick'n Conversation: Stick In-car Conversation into Places using Multi Person Finger Pointing Gestures

Kohei Matsumura
Tadashi Sakamoto
Haruo Noma
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
matsumur@acm.org {tsakamoto, noma}@mxdlab.net

Yasuyuki Sumi
Future University Hakodate
116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan
sumi@acm.org

AutomotiveUI '15 Adjunct, September 01-03, 2015, Nottingham, United Kingdom
ACM 978-1-4503-3858-5/15/09.
http://dx.doi.org/10.1145/2809730.2809735

Abstract
In-car conversations are transient and cannot be shared with others. We propose a method to stick in-car conversation into places by the recognition of multiple peoples' finger pointing gestures. We use a motion capture device to detect finger pointing gestures and their directions. Moreover, we identify a point of interest by calculating the intersection of the finger pointing gestures. Furthermore, we investigate the accuracy of the proposed method and assess user experience using an in-house car journey simulator and discuss possibilities and limitations of the proposed method.

Author Keywords
in-car conversation; multi-modal; finger pointing; knowledge sharing

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

Introduction

We propose a method to stick in-car conversation into places by the recognition of multiple peoples' finger pointing gestures. People frequently have conversations in cars. In such conversations, people often talk about points of interest they have just passed. We believe that in-car conversations are worth sharing with others because they frequently relate to a particular situation. In-car conversation topics change according to the situation. The situation can include season, time, weather, and the background knowledge of the participants in the conversations. For example, in winter, the conversations might include an observation that "this road is really slippery." However, the problem with in-car conversations is that they are transient and cannot be shared with others. Therefore, we propose a method to stick in-car conversation into places by the recognition of multiple peoples' finger pointing gestures.

In our previous studies, we proposed an in-car conversation sharing system and formulated requirements to realize the system [1, 2]. In this study, we attempt to associate an in-car conversation with a specific location such as a store or restaurant. For this purpose, we use a motion capture device to detect peoples' finger pointing gestures and their directions. We identify a target point of interest by calculating the cross point of the finger pointing gestures.

We investigate the accuracy of the proposed method and assess user experience using a car journey simulator. Moreover, we discuss possibilities and limitations of the proposed approach in view of the investigation results.

Related Work

Rümelin et al. proposed a method to identify a distant object by finger pointing [3]. They used a motion capture device to determine positions of a person's body parts. They assumed the pointing object (i.e., the point of interest) as the intersection of a face that distant with 2 m ahead of the car and calculated a vector between the person's head and hand positions. In this study, we also use a motion capture device to determine positions of a person's body parts; however, we estimate an object by calculating the intersection of multiple peoples' finger pointing gestures. This enables us to recognize the point of interest more precisely.

Numerous studies on gestural interaction have been published. Mistry and Maes introduced "Sixth Sense," a wearable device that enables interaction with projected information using gestures [4]. Their system recognizes gestures using a wearable camera and computer vision technologies. Relative to automotive user interfaces, several studies have examined gestural interactions. Ohn-bar et al. proposed a hand-gesture-based user interface for in-car entertainment [5]. In addition, Riener et al. investigated how a passenger uses space for in-car gestural interactions [6].

In-car Conversations with Finger Pointing Gestures

The results of our previous study showed that specific locations or an area could act as triggers that prompt occupants to pass on knowledge [1]. This implies that a location is a key feature of in-car conversations. The findings also raised the possibility of using finger pointing and joint attention for automatic annotation because in our experience, finger pointing and joint attention occur while a speaker discusses the target location. Thus, we assume that finger pointing detection is a key technique for annotating in-car conversations because it appears to provide indicators of objects mentioned in the conversation. During an in-car conversation, finger pointing occurs when a speaker talks about a location close to his/her car, e.g., ”this (with finger pointing) building is….” Annotating with finger pointing is possible using a motion capture system.

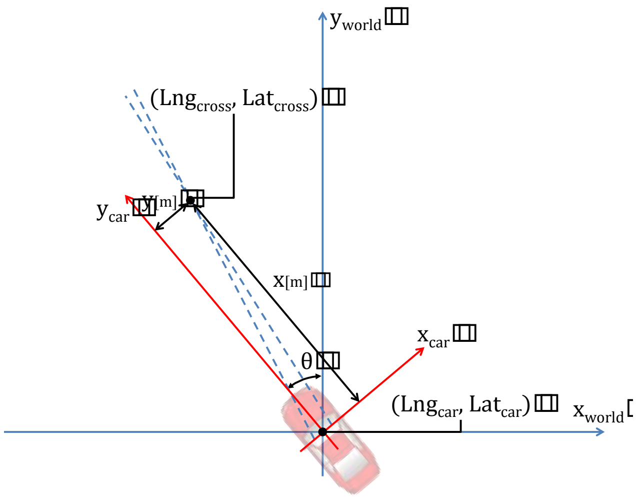

Figure 1: Overview of proposed method. We calculate the vector between head and hand positions of each passenger. We assume that the intersection of the vectors is the target point of interest. To map the world coordinates, we use car position (*Latitude and Longitude*) and direction (*theta*).

To determine the point of interest, we employ the intersection of multiple peoples’ finger pointing gestures using a motion capture device. Figure 1 shows an overview of the proposed method. We calculate vectors between the head and hand positions of each passenger. Because the intersection does not contain world information, we must translate the intersection from local coordinates to world coordinates. To perform this translation, we use the position (*Latitude and Longitude*) and direction (*theta*) of a car. In this study, we assume that the intersection is a two-dimensional field. We calculate the intersection of two passengers’ finger pointing gestures in world coordinates using Equation (2). The car’s position (Lng_{car} *and* Lat_{car}) can be obtained using the GPS and direction (*theta*) can be obtained using a compass.

$$\mu = \frac{360}{2\pi r}, \tag{1}$$

where r is the radius of the Earth.

$$\begin{pmatrix} Lng_{cross} \\ Lat_{cross} \\ 1 \end{pmatrix} = \begin{pmatrix} \mu cos\theta & -\mu sin\theta & Lng_{car} \\ \mu sin\theta & \mu cos\theta & Lat_{car} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{2}$$

Investigation

To evaluate the proposed method, we developed a system using the Microsoft Kinect v2. We tested the accuracy of the developed system and performed a user study.

Implementation

We developed a system to recognize a common point of interest for two people. We used the Microsoft Kinect v2 to determine positions of body parts of each person. We then calculated the intersection of two peoples’ finger pointing gestures using the above described method. Moreover, we evaluated the proposed system under with standalone and simulator conditions. However, it is possible to attach the system to an actual car because the Kinect v2 has a wide lens and is capable of capturing finger-pointing gestures in a small space as in a car.

Accuracy Test

We performed an experiment to confirm the accuracy of the proposed system. Eight undergraduates aged 20 to 22 participated in this experiment. We paired participants and asked them to point at a specific marker. We used a yellow tennis ball as a marker. As shown in Figure 2, we placed 11, 11, and nine markers at 3 m, 5 m, and 10 m distances from the participants, respectively. The markers were moved to one of these distances at random.

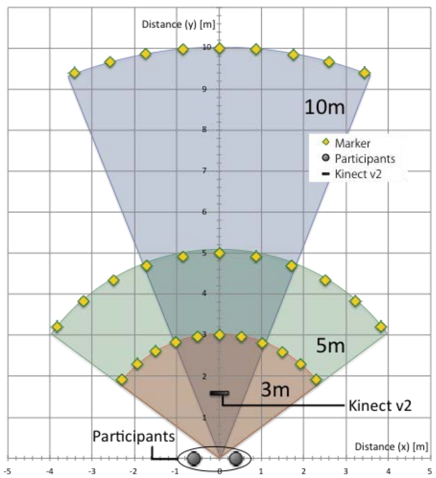

Figure 2: Experimental settings of the accuracy test

		Position of the marker			
		Left	Center	Right	Overall
Average	3m	0.44	0.39	0.38	0.40
	SEs	*0.17*	*0.14*	*0.20*	*0.17*
	5m	1.09	1.02	1.02	1.05
	SEs	*0.51*	*0.50*	*0.57*	*0.53*
	10m	2.46	2.65	2.69	2.59
	SEs	*1.19*	*1.12*	*1.19*	*1.17*

Table 1: Average distances between each calculated intersection and ground truth and their standard errors (SE) for three conditions (Unit = *meter*).

The participants were required to point at the marker with their finger for two seconds. Each participant pointed at the marker eight times for each point (248 times in total).

We collected 992 pairs of finger pointing data. Table 1 shows average distances between each calculated intersection of two participants' finger pointing gestures and the corresponding ground truth. The results indicate that errors will increase according to the distance between the target and participants. In addition, we found that the proposed system is applicable under the condition that the distance between the target and position of the car is not greater than 40 m, where we assume that the target area is 20 m^2.

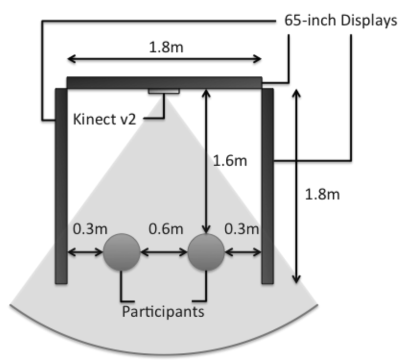

Figure 3: Experimental settings of the user study. The car journey simulator consists of three large displays and a PC.

User Experience

To investigate the applicability of the proposed system, we performed a user study with an in-house car journey simulator. The simulator consists of three 65-inch displays and a personal computer. Using the simulator, a user has a 270° field of view when driving. We installed our finger pointing detection system to the simulator to test the applicability of the proposed system in more realistic conditions than the above described accuracy test.

Figure 3 shows detailed settings of the user study. The participants sat at a distance of 1.6 m behind the front display. The side displays were at a distance of 0.3 m from the participants. The participants moved the car along a road and changed the center of the screen (i.e., rotated the car) in the simulator. The simulator can record the location (longitude and latitude), direction, and intersection of two passengers' finger pointing gestures.

Twelve people (8 males and 4 females) aged 19 to 23 participated in the user study. We paired the participants and asked them to point at landmarks two times for the conditions as follows: when a participant sees a landmark (landmark displayed on the front; front condition) and when the car passes a landmark (landmark displayed on the side; side condition). To assign three landmarks to each participant, we interviewed the participants about locations they spent their time at when they were children and selected three landmarks from these locations. If one individual is familiar with a specific location and another individual is not, then the first individual will describe the location to the second individual. We refer to this as "asymmetry of information." We assumed that this landmark selection process creates asymmetry of information between each pair of participants because they grew up in different locations.

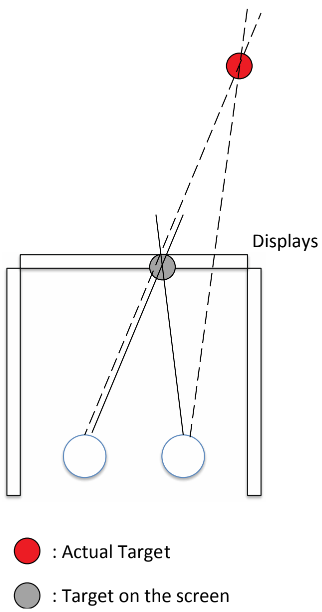

Figure 4: Participants seldom pointed the target on the display. The red circle indicates the actual target and the gray circle indicates the imaginary target displayed on the screen.

We evaluated the results by calculating distances between each pair of intersections of two participants' finger pointing gestures and a landmark. We obtained 36 data points for the front condition and 41 data points for the side condition (77 data points in total). For the front condition, 24 of the 36 data points were evaluated correctly (the intersection was at the target landmark area). However, only seven of the 41 data points were evaluated correctly for the side condition. We discuss the cause in the next section.

Findings

Errors in the front condition

Through the user study, we could not correctly estimate the target landmark for 12 of the 36 data points. Among these 12 errors, six errors were not significant (less than 10 m). We could improve accuracy by changing the position of the motion capture system (i.e., the Kinect v2). However, the other six errors could indicate a significant problem. Figure 4 shows this error situation. In this situation, the participants did not point at an actual target. They pointed at an imaginary target shown on the display. This type of error could only be found in the simulator experiment; however, it should not be ignored.

Errors in the side condition

We collected 41 data points for the side condition. However, only seven data points were measured correctly. Figure 5 shows the distribution of the estimated intersections. The distribution of the intersections is concentrated at the center of the car. Through a video investigation, we found that our estimation method was not effective for pointing at a target on the side displays. Figure 6 shows a situation wherein the participants moved their arm forward and pointed at a target using their finger. Our estimation method calculates a vector between the head and hand positions; therefore, the estimation was incorrect. To address this problem, we can calculate a vector between the hand and finger positions rather than the head and hand positions. However, this will decrease accuracy for the front condition. As a result, we must tackle a hybrid approach, i.e., while a participant is pointing at a target on the front, we can use the proposed method; however, when a participant is pointing at a target on the side, we must use an estimation algorithm. We are currently working on this estimation algorithm.

In-car conversation entertains

Through the user study, we confirmed possibilities of in-car conversations. We asked the participants to visit three places in one journey; however, five participants visited an additional place. For example, a participant was talking about a noodle restaurant and was excited about delicious noodle restaurants. They drove to a delicious noodle restaurant and kept talking about the restaurant. In addition, some participants drove to a related place after they introduced a landmark. For example, a participant went to a kindergarten school after he/she introduced his/her elementary school.

Use of characteristics of scenes

We found that the participants often used a characteristic word, such as the color, form, or name of a building, to introduce places. If we can use this information to filter targets, we can estimate the targets more accurately.

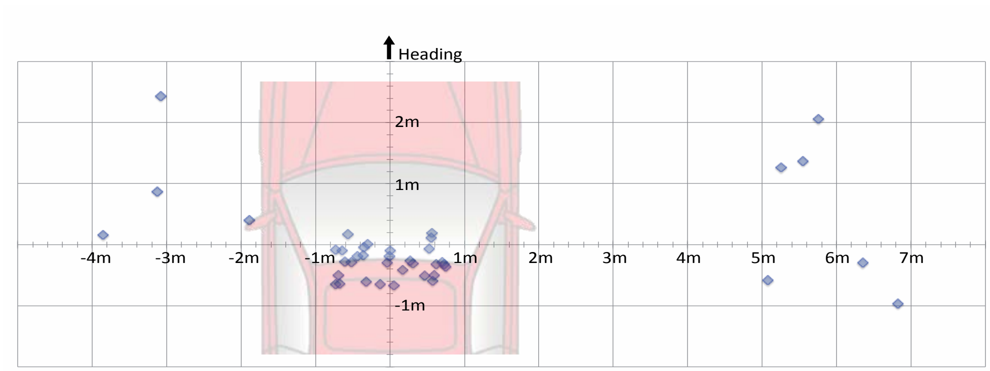

Figure 5: Distribution of estimated intersections for the side condition. Each blue square indicates the position of an intersection.

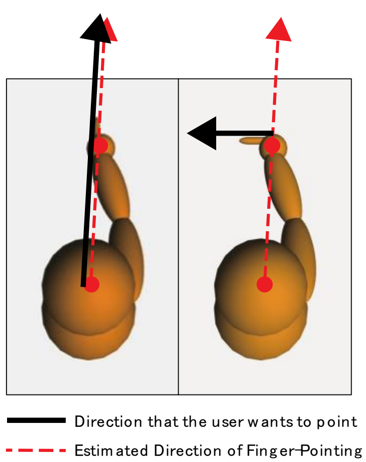

Figure 6: Misestimated situation while pointing at a target displayed on side displays.

Concluding Remarks

This paper proposed a method to stick in-car conversation into places by the recognition of multiple peoples' finger pointing gestures. To estimate the target location, we calculate the intersection of their finger pointing gestures by obtaining a vector between the head and hand positions. We investigated the accuracy of the proposed method and assessed user experience using an in-house car journey simulator. Through the investigation, we confirmed the applicability of the proposed method and identified some limitations. In future, we plan to incorporate an estimation algorithm using characteristics of target buildings to improve accuracy.

Acknowledgements

This work was supported in part by Microsoft Research 10th collaborative research program.

References

1. K. Matsumura and Y. Sumi. 2014. What Are You Talking About While Driving?: An Analysis of In-car Conversations Aimed at Conversation Sharing. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '14). ACM, New York, NY, USA, Article 34, 8 pages. DOI=10.1145/2667317.2667417
2. K. Matsumura and Y. Sumi. 2014. CarCast: a framework for situated in-car conversation sharing. In *Proceedings of the 5th Augmented Human International Conference* (AH '14). ACM, New York, NY, USA, Article 17, 4 pages. DOI=10.1145/2582051.2582068
3. S. Rümelin, C. Marouane, and A. Butz. 2013. Free-hand pointing for identification and interaction with distant objects. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '13). ACM, New York, NY, USA, 40-47. DOI=10.1145/2516540.2516556
4. P. Mistry and P. Maes. 2009. SixthSense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches* (SIGGRAPH ASIA '09). ACM, New York, NY, USA, Article 11, 1 pages.
5. E. Ohn-Bar, C. Tran, and M. Trivedi. 2012. Hand gesture-based visual user interface for infotainment. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '12). ACM, New York, NY, USA, 111-115. DOI=10.1145/2390256.2390274
6. A. Riener, et al. 2013. Standardization of the in-car gesture interaction space. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '13). ACM, New York, NY, USA, 14-21. DOI=10.1145/2516540.2516544